

Evaluating Google Machine Translation in Translating Arabic News Articles Headlines into English

Rawan Khalfan Al Nadabi & Dr. Hassan Obaid Alfadly

Department of Foreign Languages, College of Arts and Sciences, University of Nizwa,
Sultanate of Oman

Abstract

This study evaluates the performance of Google Translate (GT) in translating Arabic news headlines into English, focusing on clarity, accuracy, and stylistic quality. Headlines were collected from reputable news agencies, including Al Jazeera, BBC, Oman Daily Observer, Al Watan, and Al Arabiya. Ten headlines were translated using GT and compared with human translations, which were then evaluated by eight professional translators through a structured grid-based survey. Results indicate that GT performed well in clarity (95%) and accuracy (93.7%) but scored lower in style (90.6%). While GT demonstrates remarkable progress in machine translation, the findings highlight its limitations in maintaining stylistic nuances in complex linguistic contexts such as media translation. This underscores the importance of human oversight and the potential for hybrid approaches combining machine and human translation for optimal results.

Keywords: Machine Translation, Google Translate, Arabic-English Translation, News Headlines, Linguistic Style

1. Introduction

Globalization, which is resembled by the world wide web, social media, and international news, has demanded language barriers broken by quick translation production of machine translation (MT). According to Hutchins (1986, p: 15 as cited in Kadhim et al., 2013), machine translation is an application of computers, to translate one natural language into another. Machine translation has been known in the past decades as a computational linguistic phenomenon because of the rapid growth of technology and information that hasn't left any scientific and linguistics field without its intervention. The existence of MT and its swift development made it a worthwhile subject for researchers, commercial developers, and users.

1.1. Problem Statement

The massive use of machine translation especially Google Translate, which has been proven to be the most powerful and accurate of the readily available machine translation tool due to its reliance on a huge database compared to other machine translation systems, led to the belief that Google Translate has a high rate of accuracy exceeding other machine translation because of its constant upload of documents worldwide and with the function of users input in "suggest an edit" the improvement of the accuracy of Google translate continues according to (Tongpoon-Patanasorn & Griffith, 2020), this triggered the curiosity towards News headlines and the performance of Google Translate taking into consideration that headlines are complicated and special in terms of grammar and style. In addition, the Arabic language is known for its rich morphological and different word order, further complicating the translation process (Alqudsi et al., 2014). This raises concerns about the tool's performance in such contexts. In this research, I will evaluate Google Machine Translation

in translating Arabic news headlines into English, this research aims to contribute to the body of literature on machine translation and Arabic language processing, while also providing insights into the limitations and potential improvements for such translation tools.

1. Research Objectives

To examine the quality of Google Machine Translation in translating Arabic news article headlines into English.

2. Research Questions

How is the quality of Google Machine Translation in translating Arabic News article headlines into English?

1.2. Theoretical Framework

As I am assessing the quality of Google Machine Translation, in this research, the raw outputs of Machine Translation are going to be assessed and evaluated based on the criteria of Hutchins and Somers (1992).

Hutchins and Somers (1992) criteria focus on the translated text's intelligibility, fidelity, and style. According to them, these criteria are the most obvious translation quality test.

Hutchins and Somers defined each criterion (1992: 163), as follows:
Intelligibility, referred to as clarity, reflects the reader's understanding of the text, which means that the translated text is coherent. According to Arnold et al (1994: 169, as cited in Fiederer & O'Brien, 2009) grammatical errors, untranslated words, and mistranslated words affect intelligibility, considered the traditional way of assessing the translation quality. Fidelity, or the equivalent accuracy, means the approximate transference of the intended meaning and information of the source text into the target translated text, on other words is the preservation of the source language meaning in the translated text (ibid 1994: 171, as cited in Fiederer & O'Brien, 2009). Arnold et al (1994: 169, as cited in Fiederer & O'Brien, 2009), declared that intelligibility and fidelity correlate; the higher the intelligibility the higher the fidelity. The intelligibility and fidelity are the most important metrics of machine translation evaluation by humans as pointed out by Roturier (2006: 83). As for style, Hutchins and Somers (1992: 163), defined it as the fluence use of the language in the translation texts for its relatively appropriate content and intention. T.C Halliday, in Van Slype's 1979, as cited in Dabbadie et al 2002: 14, proposed a 4-point scale for the intelligibility measurement. To Halliday, comprehensibility and intelligibility are synonyms. Style is rarely found in the evaluation of machine translation quality assessment, as it is considered a major factor in rating a human translation, however, Hutchins and Somers (1992: 173), disagreed by highlighting that translators' instinct is to assess MT translation to the quality predicted by a human translation, therefore stylistic qualities are as important as the other two criteria. The use of Hutchins and Somers (1992) criteria in this research because it is considered a basic principle of machine translation evaluation as the researcher focuses on the quality of the Google Translate translation of Arabic headlines into English

with the comparison to human translation of the same headline. This criterion is easier for evaluators to judge the quality as it precise and basic metric.

2. Literature Review

The high translation quality produced by machine translation is due to the rapid growth in information and technology. Increased on the other hand the users' demand of using machine translation. This draws the researchers' attention to the quality evaluation of the machine translation production. The evolution and advancement of machine translation pave the way for an in-depth understanding of employing a computer to perform an automated translation. The literature review highlights the development of Google Translate over the years, and the translation quality production enhancements made it the focus of this study. With the development of machine translation outcome quality, the quality assessment approaches changed by researchers evaluating machine translation outcomes.

2.1. Evolution and Advancements in Machine Translation

In 1950, the technology of machine translation existed. The idea of employing a digital computer to perform an automated translation was first discussed in 1955 as cited in Prates et al., 2020. Since 1950, automated translation received the researchers' attention in computational linguistic research. (Prates et al., 2020). Hovy et al. (2002) pointed out that Machine Translation is a worthwhile subject for researchers, software developers, and users. According to Hutchins and Somers (1992, p. 3), machine translation is a computerised system that produces translation from one language into another with or without human assistance.

Prates et al (2020), highlighted that the utilization of Machine Translation (MT) branded the thought to the majority that automated translation output quality is equal to human translation production. It is a justifiable thought as machine translation over the past years has not been seen as an aid but rather a replacement for human translation, therefore, a massive increase in research studying the quality of machine translation was focused on tackling the shortage and errors made by the MT. Much of the research helped overcome the MT shortage, which was mostly beneficial to the software developers and resulted in the development of MT software.

Ghasemi & Hashemian, (2016) identified that machine translation software and applications are in rapid and continuous growth, different types of machine translation were subject to assessment as determined by Ghasemi & Hashemian, (2016). Such as statistical, rule-based, and recent development of machine translation is neural machine translation.

2.2. Google Translate and MT Software

It has been shared that Google Translate machine translation is one of the largest, free and available existing translation tools for over 200 million users daily (Shankland, 2017, as cited in Prates et al., 2020), Which makes it the most reliable and trusted website for the process of translation and makes it a better machine translation for this research. Its establishment owed to a mistranslated email by a licensed program company software

Google used in 2004, which caused the frustration of Sergey Brin, co-founder of Google Translate, Le et al., (2016) highlighted. Google Translate was launched two years later using phrase-based machine translation software, breaking the input phrases and translating sentences and phrases separately. In 2014, Google started using Statistical Machine Translation, which is based on looking for the best translation pattern of human translation stored in the model, depending on bilingual text corpora. In the same year, Yamada (2019), conducted a study on the Google Translate statistical model, where he studied the post-edit performance of college language learners. This research hypothesized that using the statistical model would reduce the level of post-editing effort of nonprofessional translators, yet the results of the research reflected that the use of students' post-edit effort didn't match the professional standards. The researcher explained the result due to the limitation of the statistical model at that time. Since 2014, Google Translate made an initiative called the Community Initiative, where Google received inputs and comments on translation for its users to uplift the production of translation produced by Google Translate to be double-checked for quality assurance. (Kelman, 2014 as cited in Prates et al., 2020).

In September 2016, the Google team announced the new software model that Google is presently using, neural machine translation, which came to address traditional statistical machine shortcomings. The neural machine translation is considered a subset of AI which empowers its strength in the ability of the machine to learn directly by the end-to-end process, in which the input text is mapped to output text association (Kisilu et al., 2024). It also effectively copes with long inputs due to attention mechanisms accompanied with. Add to the strength of neural machine translation is that it avoids choosing the brittle design of traditional machine translation. Wu et al. (2016) conducted research that studied the quality of NMT in comparison with SMT Google Translate used. The result of the research showed that the translation errors were reduced to 60% compared to statistical machine translation. The result of this study is expected due to the dependence of SMT on bilingual text corpora that limits the translation process of the software while NMT is a learning software that adapts to the users' input and provides better translations.

However, three inherent weaknesses of NMT affect its accuracy, the ineffectiveness in dealing with seldom used words, training, and interference slower, and the untranslated words in the source text. (Wu et al., 2016).

Kisilu et al. (2024) studied the accuracy of Google Translate when translating English-Kiswahili and Kiswahili-English Newspaper Headlines. The study used 50 random headlines retrieved from two different news agencies. The headlines were translated by Google Translate and human translators. The translations of both Google Translate and Human Translate were compared using the qualitative and quantitative research methodology. The result revealed that human translation is more accurate than Google translation although some human translations were similar to the production of Google Translate. Google was able to translate 56% of the 50 headlines accurately.

2.3. Machine Translation Assessment

Machine translation software changed over the years as well as the researchers' approaches, it started with word-by-word analysis, as a direct approach. And advanced to statistical and rule-based approaches. (Kadhim et al., 2013) The evaluation of machine translation in the early years demands the involvement of a human evaluator in the process. Most researchers are interested in studying machine translation output instead of post-editing it. However, human evaluation has its disadvantages, for its subjectivity, cost, and time-consuming, which led to the development of automated evaluation metrics. (Federer & O'Brien, 2009).

Hutchins and Somers (1992: 163), suggested the most obvious tests evaluating machine translation, which are intelligibility, fidelity, and style. Arnold et al (1994 as cited in Fiederer & O'Brien, 2009) commented that intelligibility is a traditional way of assessing quality because it reflects user quality judgment, the more the user understands the higher the score is. He claims intelligibility correlates with accuracy. Slype's Study (1979 as cited in Fiederer & O'Brien, 2009), evaluated the quality of machine translation for the European Commission, T.C Halliday proposed intelligibility as a measurement for the quality evaluation on a 4-point scale. To Dabbadie et al (2002 as cited in Fiederer & O'Brien, 2009) intelligibility is the most important measurement in evaluating the MT quality. Roturie, (2006) agreed by pointing out that intelligibility and fidelity are the most frequent metric in the human evaluation of MT. There is a notable shortage of scholarly literature addressing the quality of machine translation specifically for Arabic-English language pairs. Despite the growing interest in natural language processing and translation technologies, the evaluation of machine translation systems in the context of these languages remains underexplored, highlighting a significant research gap.

3. Research Methodology

3.1. Sample

The sample used in this research is news headlines collected from famous news agencies such as Al Jazeera Arabic-English, BBC Arabic- English, Oman Daily Observer, Al Watan, and Al Arabia Arabic-English. The headlines collected are from Arabic language source texts. The Arabic language was chosen to be the source text for its unique features as a language and for the intended purpose of the researcher's subjective language for the MT experiment as it is known for its complex morphology. It is also to assess Google Machine Translation performance in the language. For an Arabic headline to be chosen, the researcher made it necessary to have a human translation of the same headline. Therefore, ten headlines were chosen for this study. The selection of ten Arabic headlines was purposefully made to align with the study's objectives, ensuring they provide a focused yet diverse representation of linguistic and thematic features relevant to Arabic news reporting. These headlines were chosen for their richness in linguistic elements, cultural references, and relevance to the research topic, offering depth and breadth in analysis. While limited in number, they are sufficient to achieve data saturation, where recurring patterns and insights emerge without requiring additional data. Headlines, though brief, are dense with linguistic and cultural nuances, making this manageable sample size appropriate for detailed qualitative within the

study's scope and practical constraints. Additionally, this selection serves as a foundational step, offering preliminary insights that can guide broader studies with larger datasets.

3.2. Instrument

The instrument used to collect the data is a web-based survey, Google Forms. Google Forms was chosen because of its friendly user base, which easily obtains the users' responses. It could be disseminated and answered by individuals anywhere and anytime, it is very accessible and doesn't need special software to operate. Google Forms provides different options for how to design your survey, for example, you can design the response as a multiple-choice question, scale, check boxes, text, grid, and others. In this research, the grid option will be sufficient for the desired goal.

3.3. Data Collection

There are two main collections of data. The first collection consists of Arabic news headlines. These headlines were gathered from Arabic news agencies, which are either based in Arabic-speaking countries or have dedicated Arabic pages. Finding accurate human translations for the headlines within the same agency was challenging, especially for political headlines. Each agency often uses different headlines in English compared to their Arabic versions. Each agency with its two different pages was targeting the audience of each language separately most of the time. It is only the political topics that have quoted headlines from the speaker that are preserved the same in shape and purpose. However, some translations of the Arabic headlines of one news agency were found on other engines' pages with the same structure and purpose some agencies share the same headlines in both languages on the same topic.

The other collection of the data was through survey dissemination to professional translators, who practice translation as a job in newspapers agencies in Oman. The survey was sent through email with general instructions to guide them in answering the survey. I utilized the same survey questions as those employed in the research by Kadhim et al. (2013), these questions were designed based on the assessment criteria outlined by Hutchins and Somers (1992), ensuring consistency and alignment with established evaluation frameworks.

3.4. Data Analysis

The data were analyzed based on the criteria of Hutchins and Somers (1992): intelligibility, fidelity, and style, which are considered the basic principles when evaluating the quality of machine translation. In this study, evaluators are translators of English newspaper agencies in Oman. Therefore, this criterion is best suited for this study to evaluate only headlines. For this study and to make sure that evaluators are more familiar with the terms used in evaluating, I have used the synonyms of those criteria. For intelligibility, clarity is equivalent, for fidelity, the term accuracy is a synonym, yet style remains the same because, as I assume is a well-known term for evaluators. In the survey, the evaluators were guided by questions for each criterion which they needed to answer on a scale of 1-4. The evaluators were asked to assess the clarity of the MT output of a headline in comparison with an HT of

the same title. They were asked the following question: how easily can you understand Google Machine Translation in comparison to the human version of the same title?

Second, in the order of assessment comes accuracy. A translation can be understood but doesn't mean it reflects the intended meaning. On a scale of 1-4, the evaluators were asked to answer the following question, to what extent the information of the source headline contains the same information in the target headline in comparison with the human translation version of the same title?

Hutchins and Somers (1992) justify the use of style as a criterion in assessing machine translation by giving the aim of this research to assess machine translation in comparison to human translation. The evaluators were asked on a scale of 1-4 to answer, does Google Machine Translation sound natural and idiomatic in comparison to the human translation of the same source headline?

On a scale of 1-4, for clarity one implies that the translation is not understandable, two means that the translation is partially understanding of the translation, three means that the translation is mostly understandable, and four means that the translation is fully understandable. The same scores resemble the same tendency of the criteria studied based on the question.

The researcher assumes that GMT would score a high clarity, accuracy, and style ratio considering Google Translate's recent advanced performance.

4. Results & Discussion

Clarity, accuracy, and style are the three main criteria the current study focused on assessing the quality of the GMT in comparison with the HT of the randomly chosen Arabic headlines. The questionnaire was disseminated to 8 translators who were asked to evaluate each headline based on the three criteria on a scale of 1-4. Taking into account that the most active newspaper agencies are only two of the number of participants who are specialists in the field and work in English newspaper agencies in Oman.

The data received were transformed into Microsoft Excel to calculate the average of each parameter of the 10 headlines, and to extract charts.

The total average of each criterion was calculated manually following the formula:	The percentage of the total average of each criterion is calculated based on the formula:
$Totle\ Av = \frac{Av1 + Av2 + Av3 + Av4 + \dots + Av10}{n}$ <p>Av_x= average score of each evaluator of the same criterion</p> <p>n= number of evaluators Av= total average</p>	$Percentage = \frac{Av \times 100}{4}$

The first criterion chosen to be evaluated from the first sight is clarity, the evaluators were asked to evaluate on a scale of 1-4, how easily can you understand GMT translation in comparison to human translation of the same headline?

On a scale of 1-4, 1 referred to not understandable, 2 referred to partially understandable, 3 referred to mostly understandable, and 4 referred to fully understandable.

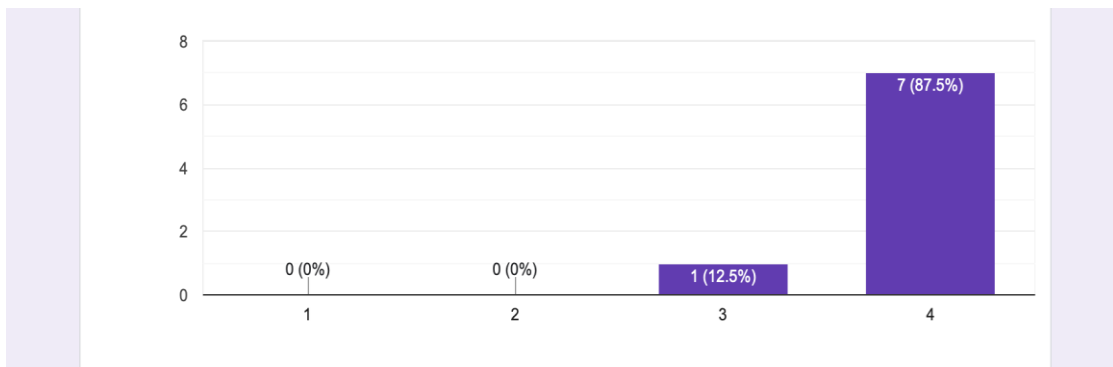


Figure 1 Clarity Evaluation of a headline

The following chart resembles every human judgment of each headline.

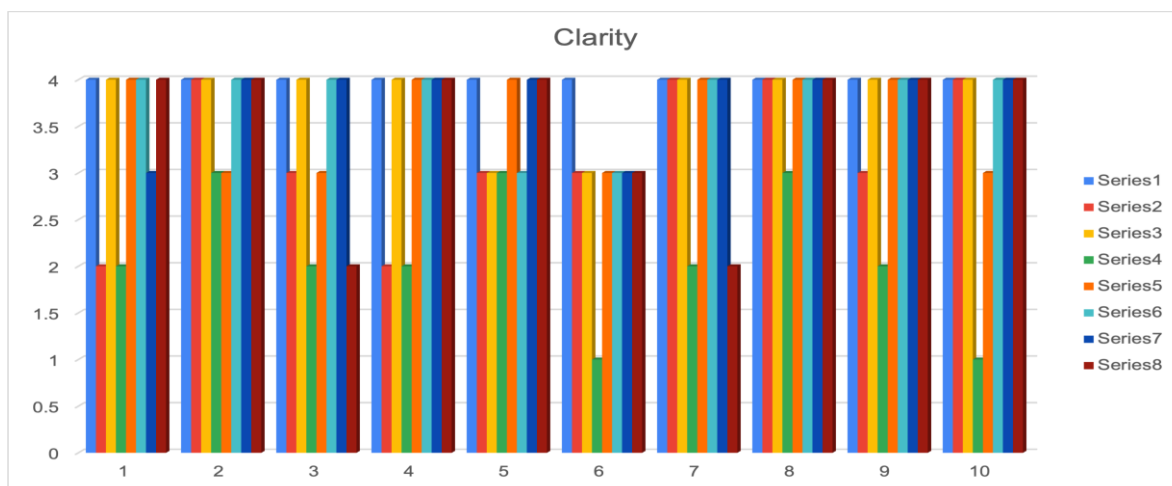


Figure 2 Clarity Evaluation Chart

In this chart, the series resembles the eight evaluators, while the X axis reflects the headlines which are 10 headlines in this study. The Y axis resembles the scale outputs 1 to 4. The chart shows that headline 8 scored the highest score, by the 8 evaluators. It also shows that evaluator four evaluated the GMT in comparison with human translation with the lowest scores in comparison with other evaluators. In headline 10 s/he evaluated the headlines 10 and 6 clarity with a score of one which reflects the translation was not understandable.

	Headline 10	Headline 6
Source	السعودية وإيران تبدآن ترتيبات إعادة فتح السفارات والقنصليات	لماذا كل هذا الغضب؟
GMT	Saudi Arabia and Iran begin arrangements to reopen embassies and consulates	why all this anger?
HT	Saudi Arabia and Iran agree to reopen embassies	What are people so angry about?

The fourth evaluator only evaluated three headlines translated by Google Translate as mostly understandable out of the ten headlines. However, headline eight scored the highest score. It scored 4 by 87.5% of the evaluators and 12.5% chose 3.

source	ماذا جرى لبولندا؟
GMT	What happened to Poland?
HT	What Happened to Poland?

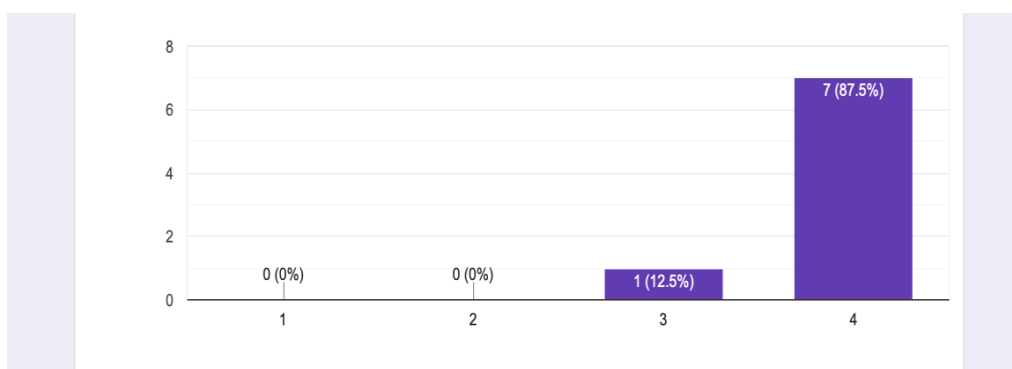


Figure 3 Headline Eight Score

Accuracy comes second in evaluation, the evaluators were asked, to what extent the information of the source headline contains the same information in the GMT target headline in comparison with the human translation. On a scale of 1-4, 1 resembled not the same information, and 4 resembled the same information. The following chart displays the evaluators' judgments of the same 10 headlines.

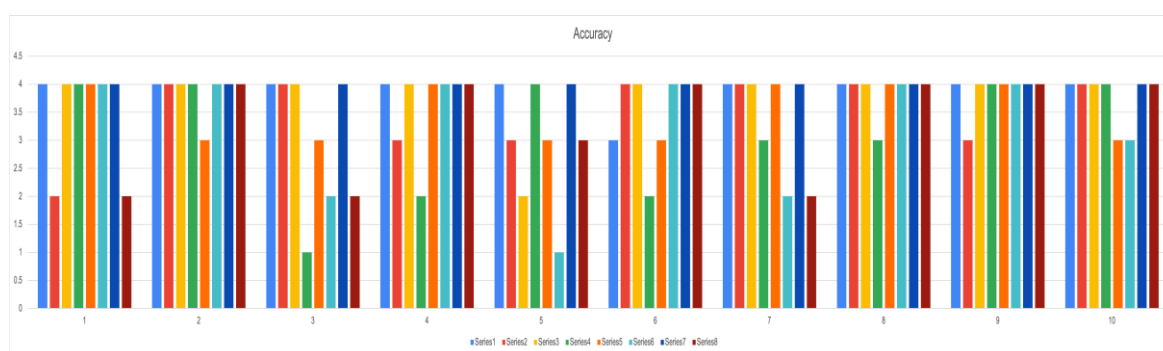


Figure 4 Accuracy Evaluation Chart

The headline with the highest score in clarity showed the highest score in accuracy which are headlines 8, 9, and 2. However, the evaluator who chose the lowest scores in clarity for headlines 10 and 6, chose a high score for accuracy for the same headlines which reflect a contradiction in the assessment of the quality of GMT translation.

The third criterion in evaluation is style. The evaluators were asked to score the translation style by answering the question, does Google Machine Translation sound natural and idiomatic compared to the human translation of the same source headline? On a scale of 1-4, 1 refers to the inappropriateness of the language, and 4 refers to the appropriateness of the language. The following chart highlights the human judgment of the translated headlines by GMT fluency and idiomatic.

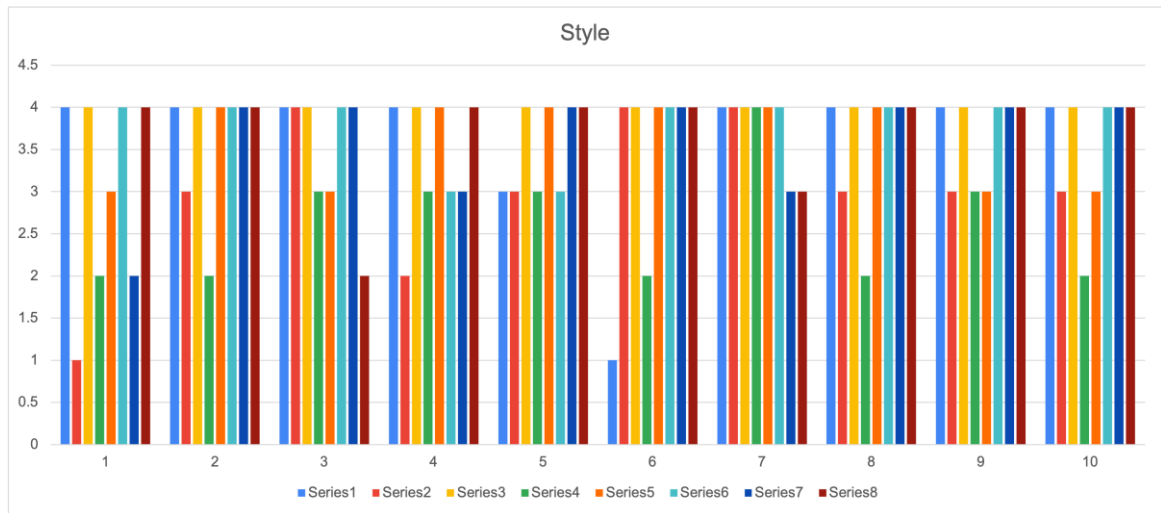


Figure 4 Style Evaluation Chart

The chart shows that headlines 7,5, and 3 score the highest score in language appropriateness. This concludes that style doesn't correlate with the accuracy and clarity of the headline translation.

	Headline 5	Headline 7	Headline 3
Source	إلى أين تتجه الاحتجاجات وهل يرضخ نتنياهو؟	اشتباكات لليلة الثانية بين الشرطة الإسرائيلية وعشرات المصلين المتحصنين داخله	محادثات رفيعة المستوى بين البلدين لأول مرة منذ سبع سنوات برعاية صينية
GMT	Where are the protests heading? Will Netanyahu acquiesce?	Clashes for the second night between the Israeli police and dozens of worshippers holed up inside.	High-level talks between the two countries for the first time in seven years under Chinese sponsorship
HT	Why is there a crisis? Will the government back down?	Israeli police storm al-Aqsa mosque for the second time on Wednesday.	Rivals Iran and Saudi Arabia hold first high-level talks in seven years.

Calculating the total average of each criterion following the mentioned formula previously. The total average of clarity is:

$$\text{Clarity total Av} = \frac{31}{8} = 3.8$$

The total average accuracy is:

$$\text{Accuracy total Av} = \frac{30}{8} = 3.7$$

The Total average of style is:

$$\text{Style total Av} = \frac{29}{8} = 3.6$$

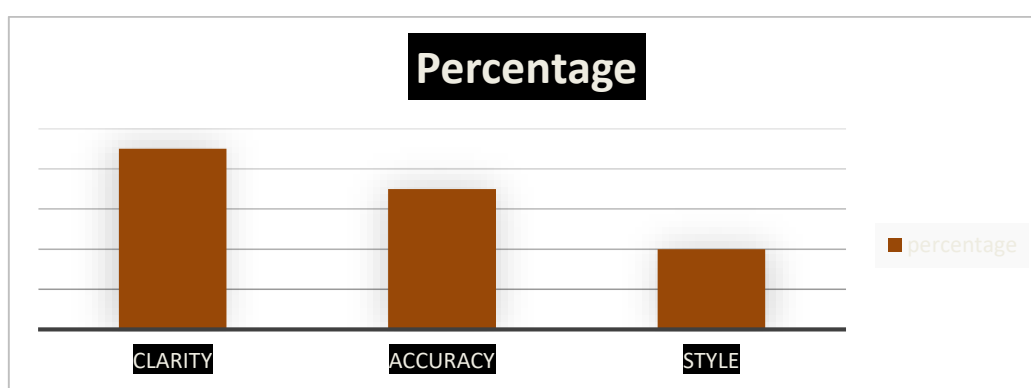


Figure 5 Overall Percentage of Each Criteria Average

Therefore, answering the research question of How is the quality of Google Machine Translation in translating Arabic News article headlines into English? Based on the analysis and statistics, the percentage of clarity is 95% based on the evaluation of translators, which means that the majority of the headlines translated by Google Translate, compared with human translators, are mostly understandable and fully understandable rated translations. This leads to the conclusion that Google Translate provided an understandable headline translation. Accuracy comes second in the highest percentage based on the statistics. Google Translate's translation of the headlines is 93.7% accurate compared to the human translation of the same headlines, indicating the correlation between clarity and accuracy. Google Translate style scored based on the evaluator's evaluation is 90.6%, which means Google was able to produce mostly appropriate translation when compared to human translation, even mostly the same human translation production of the same headline. Thus, to answer the research question, Google Translate performed excellently when translating Arabic headlines into English.

5. Conclusion

This study set out to evaluate the quality of Google Machine Translation (Google Translate) in translating Arabic news article headlines into English, focusing on key parameters such as clarity, accuracy, and style. The findings revealed that Google Translate performed

remarkably well in this context. With a clarity rate of 95%, the translations were predominantly understandable and aligned with human translation in terms of comprehensibility. The accuracy rate of 93.7% further highlights the tool's ability to produce translations that closely match human equivalents. Additionally, Google Translate achieved a style score of 90.6%, demonstrating its capability to replicate appropriate stylistic elements of human translations.

These results indicate that Google Translate is a highly effective tool for translating Arabic news headlines into English, providing reliable outputs despite the complexity of headlines and the rich morphological nature of the Arabic language. This research contributes to the understanding of machine translation performance in specialised linguistic contexts and underscores the potential for further improvement, particularly in enhancing stylistic nuances and addressing the unique challenges of Arabic translation.

5.1. Limitation and Delimitation:

The study exclusively evaluates the performance of Google Translate, disregarding other machine translation tools like Microsoft Translator, or Bing Translator. This narrows the scope and may not allow a broader comparison of different systems. The clarity, accuracy, and style scores were based on human evaluators' assessments, which may involve subjective judgments influenced by personal linguistic preferences or interpretations. News headlines are inherently unique due to their concise and stylised nature, and the findings may not generalise to other types of text, such as paragraphs, literary works, or informal texts. The number of participants in the survey limits the generalization of the conclusion, as well as the number of headlines limits the possible generalization of the study's results. The research has chosen specific types of news headlines (e.g., political, economic, or social) rather than a more diverse selection across all news categories, which could influence the findings.

References

- Kadhim, K. A., Habeeb, L. S., Sapar, A. A., Hussin, Z., & Abdullah, M. R. T. L. (2013). An Evaluation of Online Machine Translation of Arabic into English News Headlines: Implications on Students' Learning Purposes. *Turkish Online Journal of Educational Technology-TOJET*, 12(2), 39-50.
- Alqudsi, A., Omar, N., & Shaker, K. (2014). Arabic machine translation: a survey. *Artificial Intelligence Review*, 42, 549-572.
- Azer, H. S., & Aghayi, M. B. (2015). An evaluation of output quality of machine translation (Padideh Software vs. Google Translate). *Advances in Language and Literary Studies*, 6(4), 226-237.
- Bozorgian, M., & Azadmanesh, N. (2015). A survey on the subject-verb agreement in Google machine translation. *International Journal of Research Studies in Educational Technology*, 4(1).
- Dabbadie, Marianne et al (2002). "A Hands-On Study of the Reliability and Coherence of Evaluation Metrics". In Proceedings of LREC 2002 (Language Resources and Evaluation Conference). Las Palmas, Canary Islands, 27 May 2002. 816. Online at: <http://mt-archive.info/LREC-2002-Dabbadie-2.pdf> (consulted 15.08.2007).
- De Vries, E., Schoonvelde, M., & Schumacher, G. (2018). No longer lost in translation: Evidence that Google Translate works for comparative bag-of-words text applications. *Political Analysis*, 26(4), 417-430.
- Europe, Y.2018. is Neural Machine Translation Always the Best option? YAMAGATA. <https://www.yamagata-europe.com/en-gb/blog/is-neural-machine-translation-always-the-best-option>
- Fiederer, R., & O'Brien, S. (2009). Quality and machine translation: A realistic objective. *The Journal of Specialised translation*, 11(11), 52-74.
- Hutchins, John and Somers, Harold (1992). *An Introduction to Machine*
- Prates, M. O., Avelar, P. H., & Lamb, L. C. (2020). Assessing gender bias in machine translation: a case study with google translate. *Neural Computing and Applications*, 32, 6363-6381.
- Roturier, Johann (2006). An Investigation into the Impact of Controlled English Rules on the Comprehensibility, Usefulness and Acceptability of Machine-Translated Technical Documentation for French and German Users". Unpublished PhD dissertation. Dublin City University.
- Tongpoon-Patanasorn, A., & Griffith, K. (2020). Google translate and translation quality: A case of translating academic abstracts from thai to english. *Pasaa*, 60(1), 134-163.
- Translation*. London: Academic Press Limited
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... & Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation.
- Yamada, M. (2019). The impact of Google neural machine translation on post-editing by student translators. *The Journal of Specialised Translation*, 31(1), 87-106.